ARTICLE

Cloud Services and the PREMIS standard directions for digital preservation

Fabiano Ferreira de Castro [1] https://orcid.org/0000-0002-8712-2654

Rachel Cristina Vesú Alves [2] https://orcid.org/0000-0002-1649-3722

[1] Federal University of São Carlos, São Carlos, SP, Brazil / e-mail: fabianocastro.ufscar@gmail.com
[2] São Paulo Federal University, Marília, SP, Brazil / e-mail: rachelvesu@gmail.com

ABSTRACT

Introduction: Digital preservation has been one of the concerns of several areas of knowledge, especially Information Science, which, through the adoption and effective use of metadata and available metadata standards, supports the treatment of resources in digital environments. Objective: The objective of this research is to study the PREMIS metadata standard and its relationship with Cloud Services. Methodology: The Systematic Literature Review was the method that enabled the construction of a consolidated theoretical reference of the PREMIS metadata standard relationship for digital preservation in Cloud Services. Results: As a result, the discussions, actions, and initiatives found in the international scientific literature that deal with the use of metadata standards in Cloud Service are presented, as well as the cases that address PREMIS, from the perspective of Cloud Services; in addition, the study of the relationship of the semantic units of the PREMIS Data Dictionary in Cloud Services. Conclusion: It is concluded that, in view of the current technological scenario, characterized by the heterogeneity of information resources, the PREMIS Data Dictionary with its semantic units is of paramount importance for the establishment of digital preservation in Cloud Services.

KEYWORDS

Digital preservation. Cloud services. Metadata. Metadata standards. PREMIS.

Cloud Services e o padrão PREMIS rumos para a preservação digital

RESUMO

Introdução: A preservação digital tem sido uma das preocupações de diversas áreas do conhecimento, em especial da Ciência da Informação que, por meio da adoção e do uso efetivo dos metadados e dos padrões de metadados disponíveis oferece respaldo para o tratamento de recursos em ambientes digitais. Objetivo: Objetivo: O objetivo dessa pesquisa consiste no estudo do padrão de metadados PREMIS e sua relação com os *Cloud Services*. A Revisão Sistemática da Literatura foi o método que possibilitou a construção de um referencial teórico consolidado da relação do padrão de metadados PREMIS para a preservação digital em *Cloud Services*. Resultados: Como resultados, são apresentadas as discussões, as ações e as iniciativas encontradas na literatura científica internacional que versam sobre o uso de padrões de metadados em *Cloud Service*, bem como os casos que abordam o PREMIS, na perspectiva dos *Cloud Services*; além disso, o estudo da relação das unidades semânticas do Dicionário de Dados PREMIS em *Cloud Services*. Conclusão: Conclui-se que, diante do cenário tecnológico vigente, caracterizado pela heterogeneidade de recursos informacionais, o Dicionário de Dados PREMIS com suas

| 1

unidades semânticas apresenta importância capital para o estabelecimento da preservação dig ital em *Cloud Services*.

JITA: JH. Digital preservation

| 2

1 INTRODUCTION

Cloud services are increasingly widespread in society and the increasing use of this type of technology is interconnected, among other aspects, with the context of preserving the various types of objects located in digital environments, providing storage and access to content in the long term.

Considering the heterogeneity of data and information stored in different structures and in digital environments, we seek to reflect to what extent metadata and metadata standards can contribute to digital preservation in Cloud Services.

According to Tauil (2018), studies that explicitly address the relationship of metadata and metadata standards, as well as their functions and applications in the context of digital preservation in Cloud Services, have been little explored in research agendas in the field of Information Science.

In this sense, it is highlighted that Cloud Services, as well as actions aimed at digital preservation, will only be defined based on the effective use of metadata and available metadata standards, which will ensure the consistency of data and information in digital environments, for their long-term retrieval.

Thus, this research seeks to study the actions and initiatives identified in the international scientific literature that address the Data Dictionary for Preservation Metadata (PREMIS) digital preservation metadata standard, in order to trace its relationship with Cloud Services.

Therefore, through an exploratory and descriptive study supported by the Systematic Literature Review, it was possible to understand the theoretical and conceptual aspects of the PREMIS standard and its relation with Cloud Services.

This article, in addition to section 1 that presents the Introduction, the context, the guiding question, the objectives and the research methodology is structured in 5 (five) sections; in section 2, Cloud Services concepts and initiatives are presented; section 3 shows the aspects of metadata and metadata standards for digital preservation; section 4 discusses the methodology used to extract information about metadata and metadata standards and connections to Cloud Services; section 5 discusses the results and findings identified in establishing the dialogue between the PREMIS standard and Cloud Services; and, finally, section 6 presents the conclusions, observations, considerations and reflections reached in this study.

2 CLOUD SERVICES AS A RESEARCH CONTEXT

The term cloud is a concept created by professionals in Information Technology (IT) and which, according to scientific literature, appears to be an abstract concept, presenting a sense of something immaterial and impalpable. For most users of these services, it is enough to simply outsource their respective personal holdings in storage structures offered by Cloud Services companies, limiting them only to abstract reflections or, even having the simple conviction that when accessing a particular Cloud Services service, the informational contents of the respective personal holdings will be easily retrieved as soon as the user wishes.

> The genesis of the term depended on multiple sources and there is no established consensus on it, however, it only became popular in 2006 when the company Amazon launched the Elastic Compute Cloud service. It is important to note that cloud computing is the result of a set of approaches, sharing many of its characteristics with

other concepts, such as: fog computing, computational grid, client-server model, pair-by-pair model, computational aggregates, among others (DUTRA; SANT'ANA; MACEDO, 2016).

However, cloud storage platforms are designed for general use and are not specially adapted to preservation needs (RABINOVICI-COHEN et al., 2011). It is in this sense that investment in the layers of data storage and infrastructure is necessary, based on the use and application of metadata and metadata standards, which will guarantee access to digital content in the long term.

Cloud Services were developed with the advantage of providing services with easy access, low cost and with guarantees of availability and scalability (GLUSHKO, 2013). For Cloud Services to support such a large amount of stored data, it is necessary that there is a physical infrastructure, which interconnects generating a complex computational system, called Cloud Computing.

There are four models of Cloud Services: the so-called private cloud (Private Cloud) - refers to an internal data center of some organization, it can be corporate, state-owned, etc. and it means that it is not available to the public; the public cloud (Public Cloud) - provides a distributed infrastructure so that only what the user actually used is paid; the Community Cloud - was created to shape the underutilized resources of user machines; finally, the Hybrid Cloud - suggests combining the implementation of existing cloud models, merging the public cloud model with the private cloud model (FRANKS, 2015; WITTEK; DARANYI, 2012; MARINOS; BRISCOE, 2009).

Rabinovici-Cohen et al. (2011, 2013), with regard to the architecture of Cloud Services, present their structure, starting from Figure 1. It is worth noting that the architecture analyzed here is a research proposal, which may subsidize actions using metadata standards, packaging and digital object management in Cloud Services. It is an initiative of a project from the European Union that lists the digital preservation metadata in Cloud Services, in the perspective of expanding the state of the art of digital preservation, focusing on cases of commercial and scientific use, such as healthcare and financial data (RABINOVICI-COHEN et al., 2013).

| 4

Figure 1. Cloud Services architecture.
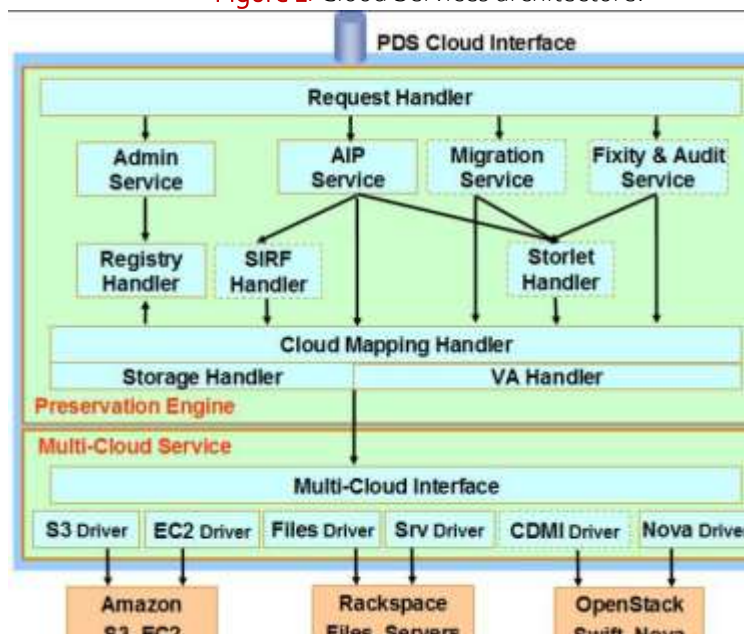
Source: Rabinovici-Cohen et al. (2013).

The Cloud Services architecture proposed by Rabinovici-Cohen et al. (2013) is represented by layers. The dotted box components are for future implementation. It is worth noting that each layer works in synergy to provide the ability to interconnect the server, storage and network components for operation. The PDS Cloud is designed as an intermediate service layer. It constitutes a broker that interconnects between Open Archival Information System (OAIS) entities and the multiple clouds. OAIS consists of a functional and information model that specifies the main criteria on which digital preservation initiatives should be based, as well as the operations to be performed in the digital environment and the information registered by metadata required for the representation of the objects/resources maintained and the long-term digital archiving (FORMENTON; GRACIOSO, 2020). In the specific context of Cloud Services, the OAIS model stands out for its relevance in the standardization of digital objects according to its established rules, that is, in the definition of a conceptual and functional model of the main information (description and content) that the Cloud Services must contemplate, as well as the actions performed by this digital environment, from the registration by metadata required for the representation and description of objects and long-term digital archiving, guaranteeing the cloud storage of the Archival Information Package (AIP).

The PDS Cloud supports logical preservation and materializes the concept of logical preservation information object in physical objects of storage in the cloud, in order to guarantee the grouping of metadata with data in the long term, helping the automation of preservation processes (RABINOVICI- COHEN et al., 2013).

PDS Cloud exposes to the client a set of preservation services based on OAIS, such as ingestion, access, exclusion and preservation actions in the AIPs of OAIS. In the internal process, it takes advantage of heterogeneous storage and calculates cloud platforms from different suppliers. AIPs can be stored in multiple clouds simultaneously to exploit different resources and pricing structures of the storage cloud and increase the survivability of the data. The PDS Cloud is divided into two main layers: Multi-Cloud Service - deals with access to a heterogeneous set of storage and cloud computing platforms. This layer is 'agnostic' to preservation; Preservation engine - provides preservation functionality for PIAs. It accepts requests through the external interface of the PDS Cloud and serves them using the multi-cloud service.

In this sense, it is possible to understand that there is an intrinsic correlation between Cloud Computing and Cloud Services, because for Cloud Services to support such a large amount of stored data it is necessary that there is a physical infrastructure, which interconnects and generates a complex computational system, called Cloud Computing (MELL; GRANCE, 2011; TAUIL, 2018).

It appears that the plurality of terminology on the topic has led to a difficulty in conceptual understanding explicitly of the term Cloud Services. For Tauil (2018), Cloud Services are digital environments that offer some type of service related to the storage of data and information, being able to outsource spaces to the user or allowing him to have access to the desired digital object, or even the availability of services infrastructure in the context of digital storage, where metadata will provide the foundation for long-term preservation.

3 Metadata for digital preservation

In Information Science, metadata is identified as a solution to promote the description and representation of resources in digital informational environments, with specific purposes

and functions, according to the knowledge community/domain.

There is a conceptual plurality in scientific literature about understanding metadata and metadata standards. In this study, the concept of Alves (2010, p 47) is adopted, which defines metadata as,

> [...] attributes that represent an entity (real-world object) in an information system. In other words, they are descriptive elements or coded referential attributes that represent characteristics that are specific to or attributed to entities [...]. Metadata standards are description structures constituted by a predetermined set of metadata (coded attributes or identifiers of an entity) methodologically constructed and standardized.

Metadata standards are established to serve certain purposes, according to the generality or specificity of different areas. Commonly, metadata standards serve the purpose of describing the information resource, however, there are standards that perform specific functions, as is the case with metadata standards for digital preservation.

To preserve digital objects, regardless of the type of storage, it is necessary to preserve not only the physical support, but also to consider several other dimensions presented by the question: logical, intellectual preservation, representation formed by metadata, and the constant monitoring of data and information that need to be preserved over time (SAYÃO, 2007).

There are five types of metadata in the literature, according to their functions: a) Administrative Metadata: "[...] used in the management and administration of collections and information resources"; b) Descriptive Metadata: "[...] used to describe and identify information about resources"; c) Preservation Metadata: "metadata related to the preservation management of collections and information resources"; d) Technical Metadata: "metadata related to the functions of the system and the behavior of metadata"; e) Metadata for Use: "metadata related to the level and type of use of collections and information resources" (GILLILAND, 2016).

According to studies by Arakaki, Alves and Santos (2019), it can be highlighted that these typologies are not excluding, since a metadata of one typology can play the role of another, for example, a preservation metadata play the role of technical metadata when describing specifications on software requirements, which is a common type of information to be registered in digital preservation.

There are two strands in the literature for categorizing preservation metadata, as highlighted by Arakaki (2019): a) preservation metadata is a subcategory of administrative metadata (RILEY, 2004, 2010; POMERANTZ, 2015; GARTNER, 2016; JOUDREY; TAYLOR; WISSER, 2018) and; b) preservation metadata is an independent category (GILLILAND, 1999, 2008, 2016; MÉNDEZ RODRÍGUEZ, 2002; SENSO; ROSA PIÑERO, 2003; HAYNES, 2018; ALVES, 2010; ALVES; SANTOS, 2013; ZENG; QIN, 2008, 2016).

Preservation metadata have the following benefits: the possibility of preserving the digital object, the collections of objects and their representations; they make it possible to register the management activities of a preservation repository; they maintain the history from changes to updates of digital objects; in addition to ensuring greater reliability in the collections of digital objects stored and that need to be accessed over time (BODLEIAN LIBRARIES, 2015).

In a cloud system, such as Cloud Services, metadata serve different functions (TAUIL, 2018; DECMAN; VINTAR, 2013; RABINOVICI-COHEN et al., 2011, 2013; BODLEIAN LIBRARIES, 2015; ASKHOJ; SUGIMOTO; NAGAMORI, 2011, 2015) and, in particular, guarantee the identity and preservation of digital documents.

| 6

Knowing that information resources are subject to continuous updates, digital preservation must be thought out and applied, including the resources and technologies available in Cloud Services. Thus, metadata, as well as metadata standards, need to be considered and applied in any initiative that has among its concerns, long-term digital preservation, as they are the ones that will guarantee the storage and persistence of data and information in Cloud Services structures, taking into account the conjunction between the technological and representation aspects necessary to the types of resources in these digital informational environments.

Thus, understanding that digital preservation is a management process that contemplates actions/strategies (structural and operational) necessary to ensure continuous access to digital objects for as long as necessary, one of the roles of metadata in Cloud Services, in addition to the storage and instantiation of data and digital objects lies in guaranteeing the reliability, authenticity and integrity of the object to be be preserved, so that its attributes, content, original features, identification and unambiguous location are maintained over time. (DIGITAL PRESERVATION COALITION, 2015; FORMENTON; GRACIOSO, 2020).

4 Methodological Procedures

The exploratory research methodology based on the Systematic Literature Review (RSL) aimed at mapping studies on the PREMIS metadata standard and its connections with Cloud Services.

The RSL consists of a bibliographic review based on the use of rigorous criteria and steps, for the prospection and retrieval of information for scientific purposes and that allows to guarantee the representativity of the retrieved documents, in addition to,

> [...] observing possible flaws in the studies carried out; know the resources needed to build a study with specific characteristics; develop studies that cover gaps in the literature bringing real contribution to a scientific field; propose innovative research themes, problems, hypotheses and methodologies; optimize available resources for the benefit of society, the scientific field, institutions and governments that subsidize science (GALVÃO; RICARTE, 2019).

To assist in the construction and development stage of RSL, the State of the Art through Systematic Review (StArt) tool was used, developed by the Software Engineering Research Laboratory (LAPES), of the Computer Science Department, at the Federal University of São Carlos (UFSCar). StArt proposes a RSL structure divided into three main stages: planning, extraction and summarization, which can be seen in Figure 2:

Figure 2. Structure of the Systematic Literature Review.
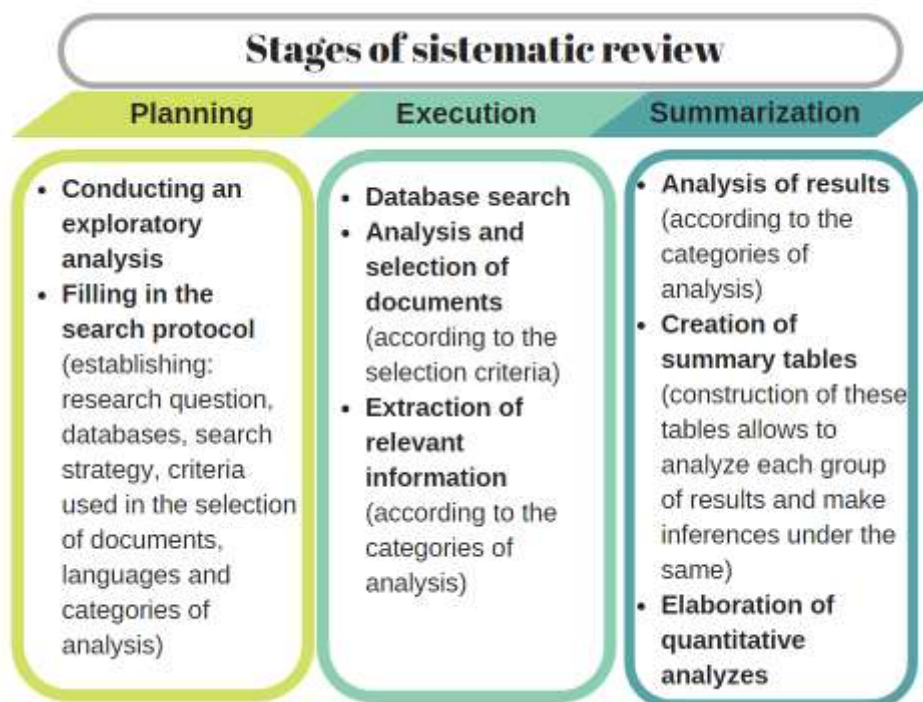
Source: Adapted from Jesus e Castro (2019, p. 47)

StArt structures a protocol in which a set of important and indispensable information is registered that will guide and conduct all stages of RSL in a categorical and systematic way. The RSL protocol comprising 18 (eighteen) elements was completed, as shown in Table 1.

Table 1. Systematic review protocol completed in StArt.

| Title	PREMIS and Preservation Metadata in Cloud Services
Description	Checking in the scientific literature the production of knowledge that addresses and analyzes the PREMIS metadata standard in Cloud Services.
Objetives	To identify and analyze the studies that deal with the relationship of PREMIS digital preservation metadata and their relationship in Cloud Services.
Purpose of the systematic review	Check in the scientific literature, specifically in the areas of Information Science and Computer Science, both nationally and internationally, publications on the PREMIS metadata standard in Cloud Services.
Intervention	Check how PREMIS digital preservation metadata studies are interconnected and their relationship in Cloud Services
Population	Digital preservation metadata used in Cloud Services.
Control	Articles published in scientific journals, interdisciplinary databases in the areas of Information Science and Computer Science and websites that include the official documents of the PREMIS standard.
Results	To identify publications that study PREMIS digital preservation metadata and its adoption in Cloud Services.
Application	Theoretical contribution for Information Science researchers in future works on digital preservation in Cloud Services.

	Relationship between the PREMIS standard and Cloud Services.
Keywords	Cloud Services; Cloud Storage; Cloud Storage Providers; Cloud Operations; Cloud Computing; Digital Preservation; Metadata Digital Preservation; Information Science Computer Science; PREMIS; PREMIS Data Dictionary for Preservation Metadata. standards
Selection criteria sources	Scientific articles, books and book chapters, theses and dissertations retrieved from the selected databases and official documents of the PREMIS Standard.
Language study	English.
Source of research methods	Systematic search using the AND and OR Booleans.
Search tool	*Computer and Information Systems* (*ProQuest*), *Web of Science* (WoS); *Information Science & Technology* (ISTA); *Library & Information Science Abstracts* (LISA) *Scopus* e; *IEEE Xplore.*
Inclusion (I) and exclusion (E) criteria	I) The main approaches of the text focus on features related to Cloud Services; I) The document presents a list of digital preservation metadata in Cloud Services; I) The document presents digital preservation strategies using preservation metadata in Cloud Services; I) The document presents digital preservation strategies using preservation metadata in Cloud Services (I); I) Only documents found in online databases; I) Any year; I) The document addresses the relation of the PREMIS standard in digital preservation strategies in Cloud Services; E) The main approaches of the text focus on characteristics related to Cloud Computing; E) The document does not present a list of digital preservation metadata in Cloud Services; E) The document includes digital preservation metadata in Cloud Services, but they are not in the areas of Information Science or Computer Science; E) The document does not address digital preservation strategies, even presenting metadata in Cloud Services, but does not relate preservation metadata with Cloud Services; E) The text is not available in full; E) The document is not in the language defined in the search.
Definition of study types	Scientific articles.
Selection of initial studies	Tauil (2018)
Extraction field information	Approaches that were used in the text in the association of digital preservation metadata studies with Cloud Services; Approaches to the text in question that are part of the scope of Information Science; Approaches to the text in question that are part of the scope of Computer Science; Text definitions on the needs to use digital preservation metadata in Cloud Services. Studies that address the PREMIS metadata standard and its adoption in Cloud Services.
Results of summarization	1. Mapping of documents that search for digital preservation metadata in Cloud Services.

| 9

	2. Focus on the documents analyzed in the Systematic Literature Review.
	3. Identification of documents that address the PREMIS metadata standard for digital preservation that relate to Cloud Services.

Source: Elaborated by authors (2020).

821 documents were identified in the RSL planning phase. 44 documents were accepted, 454 were rejected and 324 were duplicated. In the extraction phase, 11 documents were selected (25% of the total), while 28 documents were rejected (64% of the total) and 5 documents were duplicated (11%), as shown in Figure 3.

Figure 3. Documents analyzed in the extraction phase.
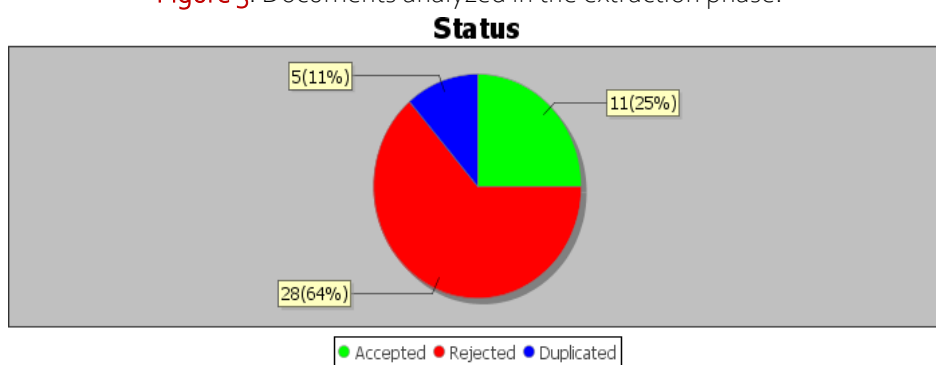
Source: Research data (2020).

In the Systematic Review of Literature, eleven documents were identified and | **10** retrieved, which were analyzed, from the thematic aspects that are directly related to digital preservation in Cloud Services, in particular, to the PREMIS standard.

5 Results and Discussion

Discussions about preservation strategies in Cloud Services started in 2011, in the areas of Information Science and Computer Science.

Considering the number of documents retrieved in Figure 3, there is a low flow of publications that deal with Cloud Services in Information Science, in a particular way, having as their main axis their relationship with the PREMIS metadata standard.

In this sense, three documents (RABINOVICI-COHEN et al., 2013; ASKHOJ; SUGIMOTO; NAGAMORI, 2011, 2015) present the attributions of the characteristics, functions and applications of the digital preservation metadata standards covering the entire path of digital objects Cloud Services and dialogue with the PREMIS standard.

Askhoj, Sugimoto and Nagamori (2011) present a model that allows the sharing of functionalities and information objects made available with other layers of cloud, based on the concepts of the OAIS reference model, focusing on the ease of sharing metadata and has the following aspects: automatic provision of metadata, reliable repository, comprehensive and standardized information packages.

Askhoj, Sugimoto and Nagamori (2011), in their study, point to the preservation layer in the sense of being more than safe storage, as it provides the types of information necessary for long-term preservation and creates information packages for the file system. Among the

metadata identified, those related to the OAIS, XML and PREMIS model stand out. The PREMIS preservation metadata standard in this initiative was divided into three subcategories of metadata:

- Metadata generated for business systems at the exact moment of creation or a registration statement: they are at least descriptive metadata for preservation that can only be provided by business systems;
- Pre-registered information metadata: statistical information must be provided to advanced systems in the cloud model, that is, information about entities registered in the system must be provided;
- Metadata of information related to events: information describing changes in digital objects and metadata during the preservation process.

In Askhoj, Sugimoto and Nagamori (2015), an ontology for cloud files is highlighted. The main metadata standard used is PREMIS, which is of paramount importance in building vocabularies for Cloud environments (creation and transfer including records of content for creating cloud systems). The ontology successively describes the components of choice, in addition to providing interoperability between the contents, creating applications and services provided by preservation metadata. The study addresses the OAIS model, specifically its characteristics for Cloud Services and relates the interaction between layers, in addition to highlighting the importance of the preservation layer (a research theme that appears in the documents analyzed in this article).

In the context of a layered model, Askhoj, Sugimoto and Nagamori (2015) state that digital objects need to be transformed into information packages, that is, they need attached metadata, as these metadata are necessary to ensure long-term preservation. The preservation layer manages breeding applications where their storage is allocated. In the cloud ontology, the semantic description is important to describe the digital file and the relationship between the metadata schemes used in the creation of the preservation metadata application, used in the preservation layer.

For each digital object allocated in the preservation layer, there are four different sources of preservation metadata: Metadata registered by the producer - consists of the metadata that was pre-registered by the producer who used the registration model; Producer-oriented metadata - supplied to the producer at the time of export; Automatic Record Preservation Metadata - created by the preservation service at the time of import; Automatic file information metadata - concerning to the treatment of system metadata with file properties, such as size, creation date and extension. (ASKHOJ; SUGIMOTO; NAGAMORI, 2015).

Finally, Rabinovici-Cohen et al. (2013) present a storage service model, Preservation Data Stores in the cloud (PDS Cloud), which seeks to maintain the ability to understand digital content (logical preservation) in the long term, adhering to the dynamic changes in preservation requirements and adapting to the evolution of technology. The Preservation Data Stores in the Cloud initiative manages to support logical preservation, being characterized as an agent that interconnects with the OAIS reference model with several existing Cloud Services. The proposed model is a service designed to work as an intermediate layer, as it interconnects between OAIS and several Cloud Services, exposing a set of preservation services based on OAIS information packages, exemplified by inclusion, access, exclusión and preservation actions.

It is worth mentioning that this initiative serves as a conceptual framework of

| **11**

RDBCI: Rev. Dig. Bibliotec e Ci. Info. / RDBCI: Dig. J. of Lib. and Info. Sci.| Campinas, SP | v.18| e021001| 2021

reference for files and providers that aim to store and preserve records in Cloud Services. Within the scope of the model developed by Rabinovici-Cohen et al. (2013), a provider only needs to bring a type of metadata that meets the requirements of the preservation service.

In relation to the advantages pointed out by the scientific literature, of using Cloud Services to guarantee the preservation of digital objects, the following stand out:

- cloud services can provide easy, automated replication to multiple locations and access to professionally managed digital storage and health verification. As a result, the bit preservation (durability) of digital information can be at least as good (or better) than it can be achieved locally;
- files can add access to dedicated tools, procedures, workflow and dedicated service contracts, customized for digital preservation requirements through specialized vendors;
- the flexibility of the cloud allows tests and pilots of providers relatively fast and low cost, for example, more options in the implantation of cloud services and, therefore, greater relevance to the files compared to previous years; in particular, private cloud or hybrid cloud implementations can address security issues about storing more sensitive material, perhaps considered inappropriate for the public cloud;
- exit strategies can be put in place to address archiving issues about provider stability and longevity or other risks of change, for example, synchronizing content between two cloud service providers or an external cloud with local internal storage; or agree to a copy of warranty maintained independently by a trusted third party (DIGITAL PRESERVATION COALITION, 2015).

Regarding the disadvantages of adopting Cloud Services for digital preservation, the following aspects are considered:

- cloud storage and service contracts need careful management over time to meet archiving needs. Data held in files should be expected to be preserved and accessible beyond the commercial life span of any current technology or service provider;
- public cloud services tend to charge each month for the capacity that has actually been consumed. As a result, it can be difficult to budget in advance or accurately predict the amount of data that is likely to be loaded, stored or downloaded (however, some vendors may charge an annual subscription based on volume);
- make sure that all legal requirements and obligations relating to the rights of third parties in the data to be stored, or on them, are met. These may be related to management, preservation or access, and may have been placed in the archives and their main organizations by their donors and financiers through contracts and agreements or through government legislation. (DIGITAL PRESERVATION COALITION, 2015).

In this sense, from the metadata standards, computational languages and reference models for digital preservation identified in the RSL in this study, the PREMIS metadata standard was chosen to analyze its relationship with Cloud Services, as according to the analysis of the identified documents, PREMIS is the most cited and highlighted metadata standard in the scientific literature to ensure digital preservation and its approximations with Cloud Services.

| 12

5.1 Study of the PREMIS metadata standard in the light of Cloud Services

The PREMIS digital preservation metadata standards, which stands for "PREservation Metadata: Implementation Strategies", is the name of a working group led by the Online Computer Library Center (OCLC) and the Research Libraries Group (RLG), between 2003 and 2005. The result of this working group is the report "PREMIS Data Dictionary for Preservation Metadata" (PREMIS Data Dictionary for Preservation Metadata), the latest version of which, 3.0, has been updated in November 2015 and it is being used in this work (CAPLAN, 2009).

The PREMIS Data Dictionary defines a set of semantic units distributed in four entities of its data model (object, event, agent and rights). These semantic units indicate which elements are necessary to perform the preservation functions in digital informational environments. They constitute the most used elements by most preservation repositories, necessary to guarantee the registration of changes that the object undergoes over time, for example, digital preservation strategies used, changes in versions and file formats of the object, its access, its authenticity, among other issues (CAPLAN, 2009).

The semantic units, also called properties, are divided into a simple data model in PREMIS (2015), which presents four entities: object, event, agents and rights. According to PREMIS (2015) they are specifically:

- Object Entity: aggregates information about a digital object maintained in a preservation repository. The object can be of the following types: an intellectual entity (a traditional or digital object such as a book, map, photo, database, etc.); a digital representation (the metadata that describes the intellectual entity); the file (named and ordered sequence of bytes known to an operating system) and the bitstream (contiguous or non-contiguous data in a file that have significant properties for preservation purposes).
- Event Entity: aggregates information about the actions that modify the objects and must be registered separately from the object.
- Agent Entity: aggregates information about agents (people, organizations or softwares) to unambiguously identify them.
- Entity Rights: aggregate information about declarations of rights to agents and legal permissions to access objects in the repository.

Each aforementioned entity presents mandatory and optional semantic units, which are made up of mandatory and optional subunits. Thus, the optional semantic unit, which has a mandatory subunit, was also selected and considered mandatory in this study.

It is worth mentioning that the hierarchical structure of the semantic units in full can be consulted in the PREMIS Data Dictionary (PREMIS, 2015). Thus, table 2 presents the semantic units considered in this research and that can potentially be applied in Cloud Services.

| 13

Table 2. Semantic units of PREMIS metadata standard entities.

1 SEMANTIC UNITS OF THE OBJECT ENTITY	
1.1 objectIdentifier	1.27 signatureValue
1.2 objectIdentifierType	1.28 signatureValidationRules
1.3 objectIdentifierValue	1.29 environmentFunction
1.4 objectCategory	1.30 environmentFunctionType
1.5 preservationLevel	1.31 environmentFunctionLevel

1.6 preservationLevelValue
1.7 objectCharacteristics
1.8 fixity
1.9 messageDigestAlgorithm
1.10 messageDigest
1.11 format
1.12 formatDesignation
1.13 formatName
1.14 formatRegistry
1.15 formatRegistryName
1.16 formatRegistryKey
1.17 inhibitors
1.18 inhibitorTyp
1.19 storage
1.20 contentLocation
1.21 contentLocationType
1.22 contentLocationValue
1.23 signatureInformation
1.25 signatureEncoding
1.26 signatureMethod

1.32 environmentDesignation
1.33 environmentName
1.34 environmentRegistry
1.35 environmentRegistryName
1.36 environmentRegistryKey
1.37 relationship
1.38 relationshipType
1.39 relationshipSubType
1.40 relatedObjectIdentifier
1.41 relatedObjectIdentifierType
1.42 relatedObjectIdentifierValue
1.43 relatedEventIdentifier
1.44 relatedEventIdentifierType
1.45 relatedEventIdentifierValue
1.46 linkingEventIdentifier
1.47 linkingEventIdentifierType
1.48 linkingEventIdentifierValue
1.49 linkingRightsStatementIdentifier
1.50 linkingRightsStatementIdentifierType
1.51 linkingRightsStatementIdentifierValue

2 SEMANTIC UNITS OF THE EVENT ENTITY

2.1 eventIdentifier
2.2 eventIdentifierType
2.3 eventIdentifierValue
2.4 eventType
2.5 eventDateTime
3.6 linkingAgentIdentifier

2.7 linkingAgentIdentifierType
2.8 linkingAgentIdentifierValue
2.9 linkingObjectIdentifier
2.10 linkingObjectIdentifierType
2.11 linkingObjectIdentifierValue

3 ENTITY SEMANTIC UNITS RIGHTS

| **| 14** |

3.1 rightsStatement
3.2 rightsStatementIdentifier
3.3 rightsStatementIdentifierType
3.4 rightsStatementIdentifierValue
3.5 rightsBasis
3.6 copyrightStatus
3.7 copyrightJurisdiction
3.8 copyrightDocumentationIdentifier
3.9 copyrightDocumentationIdentifierType
3.10 copyrightDocumentationIdentifierValue
3.11 statuteInformation
3.12 statuteJurisdiction
3.13 statuteCitation
3.14 statuteDocumentationIdentifier
3.15 statuteDocumentationIdentifierType
3.15 statuteDocumentationIdentifierValue
3.17 otherRightsInformation
3.18 otherRightsDocumentationIdentifier
3.19 otherRightsDocumentationIdentifierType

3.20 otherRightsDocumentationIdentifierValue
3.21 otherRightsBasis
3.22 otherRightsApplicableDates
3.23 startDate
3.24 endDate
3.25 otherRightsNote
3.26 rightsGranted
3.27 act
3.28 termOfGrant
3.29 startDate
3.30 termOfRestriction
3.31 startDate
3.32 linkingObjectIdentifier
3.33 linkingObjectIdentifierType
3.34 linkingObjectIdentifierValue
3.35 linkingAgentIdentifier
3.36 linkingAgentIdentifierType
3.37 linkingAgentIdentifierValue

4 SEMANTIC UNITS OF THE AGENT ENTITY

4.1 agentIdentifier
4.2 agentIdentifierType
4.3 agentIdentifierValue
4.4 linkingEventIdentifier
4.5 linkingEventIdentifierType
4.6 linkingEventIdentifierValue

4.7 linkingRightsStatementIdentifier
4.8 linkingRightsStatementIdentifierType
4.9 linkingRightsStatementIdentifierValue
4.10 inkingEnvironmentIdentifier
4.11 linkingEnvironmentIdentifierType
4.12 linkingEnvironmentIdentifierValue

The semantic units of the PREMIS Data Dictionary highlighted in this study are considered minimal for the digital preservation of objects in Cloud Services. Divided into four entities (object, event, agents and rights), the semantic units selected for this analysis represent and preserve minimal information, for example, information about the object and its characteristics; information about the actions that modify these objects; information about the people or institutions involved in the production and alteration of these objects; and information on issues of rights and access to these objects. However, the semantic units to be implemented need to be mapped to the metadata established in PREMIS XML Schema, which include the corresponding metadata expressed in the XML syntax and processable by machine.

The object inserted in a preservation repository in Cloud Services is sent in a package along with its metadata, so it is possible to find in that storage package the file's own metadata (automatically generated when the object is created); metadata from descriptive standards associated with this object; the administrative, technical and preservation metadata generated by the object's source system, in addition to the specific metadata of a preservation standard, such as the PREMIS standard, which can be included in the package of storage.

The purpose of the Self-contained Information Retention Format (SIRF) contained in the Cloud Service architecture of Figure 1 (request handler) is to function as an object storage package and its metadata, therefore it is based on the structure of the AIP package of the OAIS model and in some PREMIS standard metadata. SIRF operates as a storage package for a set of digital preservation objects, therefore, it defines the metadata of the objects that will compose the catalog of its storage structure, the relationship between the objects and their information to support the implementation of the preservation processes, for example, migration (RABINOVICI-COHEN, 2020; RABINOVICI-COHEN et al., 2011).

The characteristics of the PREMIS digital preservation metadata standard are important for ensuring interoperability between content creation applications and the digital preservation service in Cloud Services.

Thus, it is highlighted that the entire process of assigning PREMIS semantic units needs to be carried out during the entire life cycle of the digital object in Cloud Services, thus ensuring that these objects maintain their original characteristics and informations, as well as the registration of changes in the digital preservation process in Cloud Services.

6 FINAL CONSIDERATIONS

This research aimed to study the PREMIS metadata standard for digital preservation and its relationship with Cloud Services.

The Systematic Literature Review was the method used in this research, which revealed few publications that address the digital preservation metadata standards in Cloud Services in the areas of Information Science and Computer Science.

The theme of digital preservation metadata is explicitly presented in the documents of Askhoj, Sugimoto and Nagamori (2011, 2015) and Rabinovici-Cohen et al. (2013), highlighting the life cycle paths of digital objects, considering the interconnection of the assignments, characteristics, functions and applications of the preservation metadata standards in Cloud Services.

| 15

The research revealed that the PREMIS metadata standard, as well as the OAIS reference model, are the basis for the main actions related to long-term digital preservation from the perspective of Cloud Services.

The relationship between the PREMIS metadata standard and Cloud Services can be verified in the use of its semantic units, which are mandatory requirements for the structuring and storage of long-term preservation data, especially in the transfer of heterogeneous metadata schemes from other Cloud Services.

Given the current technological scenario, characterized by the heterogeneity of information resources, Cloud Services can be a viable strategic alternative for solving the constant challenges of digital preservation, however, it is a sine qua non condition that these environments adopt and apply the metadata standards of preservation effectively and appropriately and, in this context, the PREMIS standard presents itself as an interesting option.

Thus, the PREMIS Data Dictionary with its semantic units is of paramount importance for the adoption and establishment of digital preservation metadata necessary for the structuring, storage and access to data in Cloud Services and in any actions and initiatives concerned with preserving data and long-term information in the digital environment.

CRediT

| **16**

REFERENCES

ALVES, R. C. V.; SANTOS, P.L.V.A.C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. Tese (Doutorado em Ciência da Informação) – FFC, UNESP. Marília, 2010.

ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família PROV. Tese (Doutorado em Ciência da Informação) – FFC, UNESP. Marília, 2019.

ARAKAKI, F. A.; ALVES, R. C. V.; SANTOS, P. L. V. A. C. Preservação digital e proveniência: interseções entre PREMIS e o PROV. *In*: ENCONTRO NACIONAL DE

PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 10., 2019. Florianópolis. **Anais** […]. Florianópolis: ENANCIB, 2019.

ASKHOJ, J.; SUGIMOTO, S.; NAGAMORI, M. Developing an ontology for cloud-based archive systems. **International Journal of Metadata Semantics and Ontologies**, v. 10, n. 1, p. 1-11, Jan. 2015. Available on: https://dl.acm.org/doi/10.1504/IJMSO.2015.068253. Access at: 20 Apr. 2020.

ASKHOJ, J.; SUGIMOTO, S.; NAGAMORI, M. Preserving records in the cloud records. **Management Journal**, v. 21, n. 3, p. 175-187. 2011. Available on: https://www-emerald.ez31.periodicos.capes.gov.br/insight/content/doi/10.1108/09565691111186858/full/pdf?title=preserving-records-in-the-cloud. Access at: 20 Apr. 2020.

BODLEIAN Libraries. **Introduction to digital preservation**: PREMIS metadata. 2015. Available on: https://libguides.bodleian.ox.ac.uk/digitalpreservation/premis. Access at: 20 Apr. 2020.

CAPLAN, P. **Understanding PREMIS**. Washington, DC: Library of Congress Network Development and MARC Standards Office. 2009. Available on: http://www.loc.gov/standards/premis/understanding-premis.pdf. Access at: 10 May 2020.

DECMAN M.; VINTAR M. A possible solution for digital preservation of e-government: a centralised repository within a cloud computing framework. **Aslib Proceedings**, v. 65, n. 4, p. 406-424, Apr. 2013.

DIGITAL PRESERVATION COALITION. **Digital preservation handbook**. 2 nd. ed. [Reino Unido]: University Gardens, University of Glasgow, 2015. Available on: https://www.dpconline.org/handbook. Access at: 20 Apr. 2020.

DUTRA, M. L.; SANT'ANA, R. C. G.; MACEDO, D. D. J. Sublimação de dados: dos objetos físicos às nuvens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016. Salvador. **Anais** [...]. Salvador: ENANCIB, 2016.

FORMENTON, D.; GRACIOSO, L. Preservação digital: desafios, requisitos, estratégias e produção científica. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação. Campinas, v. 18, p. 1-27. 2020. DOI: 10.20396/rdbci.v18i0.8659259. Available on: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8659259. Access at : 10 dez. 2020.

FRANKS, P. C. **Government use of cloud-based long term digital preservation as a service**: an exploratory study. Granada, Spain: [s.n.], 2015. p. 371-374.

GALVÃO, M. C. B.; RICARTE, I. L. M. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion**: Filosofia da Informação, Rio de Janeiro, v. 6, n. 1, p. 57-73, set. 2019.

GARTNER, R. M. **Metadata**. New York, NY: Springer Berlin Heidelberg, 2016.

| 17

GILLILAND, A. J. Setting the Stage. *In*: BACA, M. (org.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. Available on: http://www.getty.edu/publications/intrometadata. Access at: 20 Apr. 2020.

GILLILAND, A. J. Setting the Stage. *In*: BACA, M. (org.). **Introduction to metadata**. Los Angeles: Getty Research Institute, 1999. Available on: http://www.getty.edu/publications/intrometadata. Access at: 20 Apr. 2020.

GILLILAND, A. J. Setting the Stage. *In*: BACA, M. (org.). **Introduction to metadata**. 2 ed. Los Angeles: Getty Research Institute, 2008. Available on: http://www.getty.edu/publications/intrometadata. Access at: 20 Apr. 2020.

GLUSHKO, R. J. (ed.). **The discipline of organizing**. Massachusetts, Londres: MIT Press, 2013.

HAYNES, D. **Metadata for information management and retrieval**: understanding metadata and its use. [S.l.]: Facet Publishing, 2018.

JESUS, A. F.; CASTRO, F. F. de. Dados bibliográficos para o linked data: uma revisão sistemática de literatura. **Brazilian Journal of Information Studies: Research Trends**, Marilia, v. 13, n. 1, p. 45-55. 2019.

JOUDREY, D.N.; TAYLOR, A.G.; WISSER, K.M. **The organization of information**. 4 ed. Santa Barbara, California: Libraries Unlimited, 2018. (Library and information science text series).

MARINOS, A.; BRISCOE, G. Community cloud computing. *In*: IEEE international conference on cloud computing, 1., 2009. Beijing, China. **Proceedings** […]. Heidelberg: Springer, 2009.

MELL, P.; GRANCE, T. **The NIST definition of cloud computing**: recommendations of the National Institute of Standards and Technology. Gaithersburg, Maryland: NIST, 2011.

MÉNDEZ RODRÍGUEZ, E.M. **Metadatos y recuperación de información**. Gijón, Asturias: Ediciones Trea, 2002. (Biblioteconomía y administración cultural, 66)

POMERANTZ, J. **Metadata**. Cambridge, Massachusetts; London, England: The MIT Press, 2015. (The MIT Press essential knowledge series).

PREMIS data dictionary for preservation metadata. **PREMIS version 3.0**. 2015. Available on: http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf. Access at: 20 Apr. 2020.

RABINOVICI-COHEN, S. *et al*. PDS cloud: long term digital preservation in the cloud. *In*: IEEE International conference on cloud engineering, 6., 2013. Santa Clara, California. **Proceedings** […]. Santa Clara, California: IEEE, 2013.

RABINOVICI-COHEN, S. *et al*. Towards SIRF: self-contained information retention format. *In*: Annual international conference on systems and storage, 4., 2011. Haifa, Israel. **Proceedings** […]. Haifa, Israel: ACM, 2011.

| **18**

RABINOVICI-COHEN, S. **SNIA Long Term Retention for Medical AI Applications**. 2020. Available on: https://www.snia.org/sites/default/files/SDCEMEA/2020/SIRF_Medical_AI_v2.pdf. Access at: 20 Apr. 2020.

RILEY, J. **Understanding metadata**. NISO Press: National Information Standards Organization (U.S.), 2004.

RILEY, J. **Glossary of metadata standards**. 2010. Available on: http://jennriley.com/metadatamap/seeingstandards_glossary_pamphlet.pdf. Access at: 20 Apr. 2020.

SAYÃO, L. F. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianopólis, v. 12, p. 18-47, 2007.

SENSO, J.A.; ROSA PIÑERO, A. El concepto de metadato: algo más que descripción de recursos electrónicos. **Ciência da Informação**, Brasília, v. 32, n. 2, 2003.

TAUIL, J. C. **Metadados de preservação digital em cloud services**. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos. São Carlos, 2018.

WITTEK P.; DARANYI S. Digital preservation in grids and clouds: a middleware approach. **Journal of Grid Computing**, v. 10, n. 1, p. 133-149, 2012.

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-Schuman Publishers, 2008.

ZENG, M. L.; QIN, J. **Metadata**. 2.ed. Chicago: Neal-Schuman Publishers, 2016.

| **19**

| 20

turnitin

Article submitted to the similarity system